Análisis Probabilístico de Riesgo Ambiental

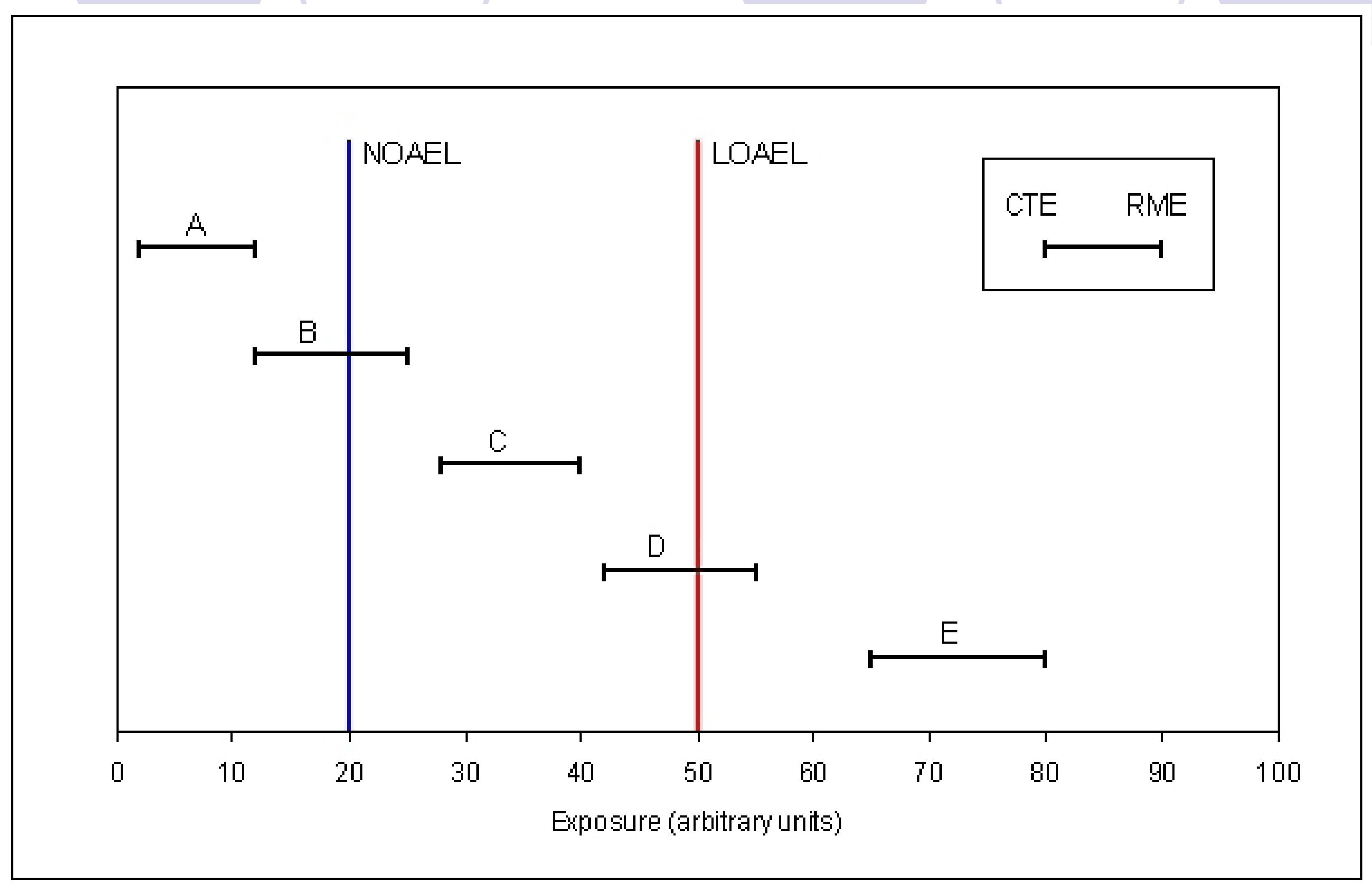
Distribuciones de Probabilidad

M.Sci.Laura Pruzzo

Motivación

- En los algoritmos de exposición y riesgo pueden combinarse multiplicativamente valores extremos
- El estimador puntual resultante podría ser poco realista
- Se ignora la definición fundamental del riesgo como la *probabilidad* de un evento adverso
- La teoría de probabilidades provee herramientas para el análisis de riesgo en el marco de dicha definición

Cuando puede ser útil el análisis probabilístico?



CTE= estimador de tendencia central; RME= máxima exposición razonable.

Fuente RAGS EPA, 1999

Cuando puede ser útil el análisis probabilístico?

- *Casos A y E: el rango de riesgo calculado mediante el método de estimación puntual está muy por debajo (A) o muy por arriba (E) del nivel crítico dado por el rango NOAEL/LOAEL*
- *El análisis probabilístico probablemente no alterará las decisiones de manejo del riesgo con respecto a la necesidad de mitigación*
- *Casos B-C-D : los estimadores puntuales de riesgo se superponen o caen dentro del rango de potencial preocupación*
- *Sugiriendo que el análisis probabilístico podría ser útil para sustentar decisiones de manejo del riesgo*

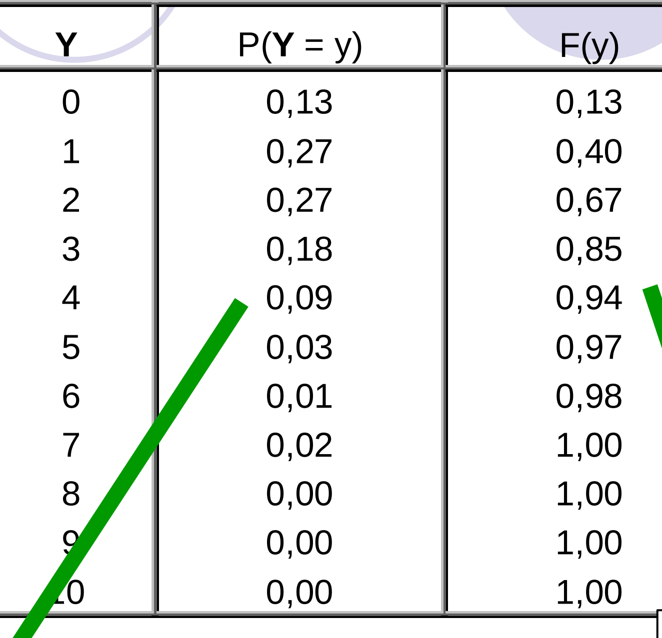
Motivación

- Deseamos representar la incertidumbre acerca de una cantidad empírica mediante una distribución de probabilidad
  - *Concentración de radón en el interior de la vivienda*
  - *Potencia carcinogénica de un químico*
- Si tenemos datos empíricos directos, podemos seleccionar una distribución y estimar sus parámetros
- Si no tenemos observaciones de la cantidad, podríamos expresar una opinión mediante una distribución de probabilidad subjetiva

Proceso aleatorio

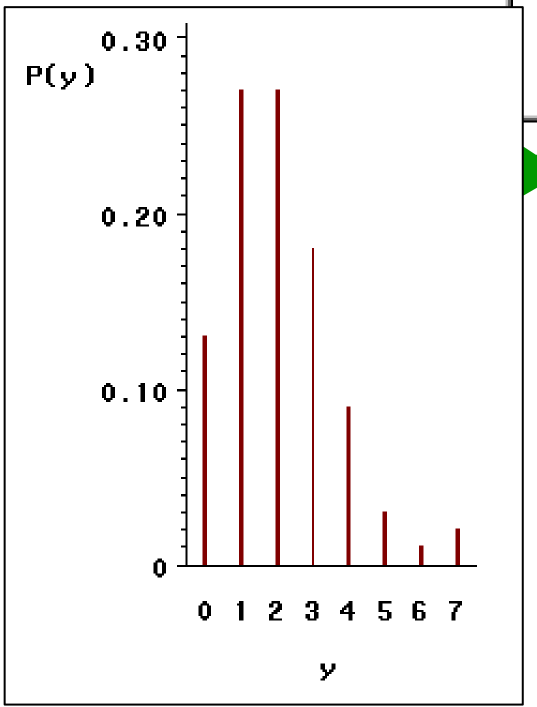
- Conjunto de posibles resultados, cada uno con una determinada chance de ocurrir
- ➔*Probabilidad* asignada a cada resultado
  - Valor de una función que asigna números reales a diversos subconjuntos de eventos
- Variable aleatoria (v.a.)
  - Variable que puede tomar un valor entre un rango, con una determinada probabilidad de ocurrencia

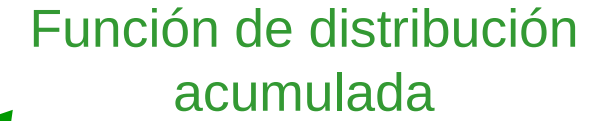
Ejemplo: Sea la v.a. **Y**=nº de pariciones en 100 hembras de búfalo.



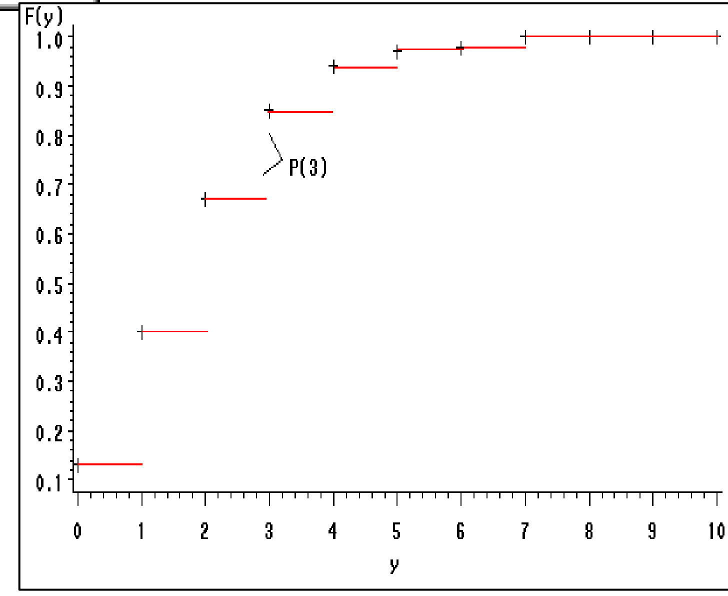
Y	P(**Y** = y)	F(y)
0	0,13	0,13
1	0,27	0,40
2	0,27	0,67
3	0,18	0,85
4	0,09	0,94
5	0,03	0,97
6	0,01	0,98
7	0,02	1,00
8	0,00	1,00
9	0,00	1,00
10	0,00	1,00

Función de probabilidad





Función de distribución acumulada



Distribuciones de probabilidad

- El rango de valores que la v.a. puede tomar y su probabilidad asociada están codificados en una función matemática llamada Función de Probabilidad de la variable
- En vez de desplegar en una tabla, se proporciona una fórmula que expresa las probabilidades por medio de una función
- Hay diferentes maneras de representar una función de probabilidad
- En Análisis de Riesgo hay muchos problemas en los que interesa conocer la probabilidad de que el valor de la v.a. sea menor o igual a algún número real *x*
- Distribución acumulada

Variables y constantes

- Variables
  - *Discretas - Probabilidad*
  - *Continuas – Función de densidad*
- Constantes: cantidades que permanecen constantes durante el proceso aleatorio
  - *Pi*
  - *Número atómico del hidrógeno*
  - *Número de días de enero*
- Parámetros de la distribución
  - *Constantes desconocidas*
  - *Se estiman a partir de los datos*
  - *Determinan la forma y posición de la distribución*

Naturaleza de la probabilidad

- Enfoque frecuentista
- La probabilidad representa la frecuencias relativas de ocurrencia de un eventos observable
- Implica repetir y medir el evento un gran número de veces y ver cuantas veces ocurre el resultado de interés
- P (evento A) = n/N
- Problema: cuando no tenemos oportunidad de repetir el ensayo un número grande de veces

Naturaleza de la probabilidad

- Probabilidad Subjetiva
- Para eventos no repetibles o mensurables
- Expresión del grado de creencia acerca de la ocurrencia de un evento incierto
- Subjetiva por naturaleza: es posible que dos personas asignen a un mismo evento, diferente probabilidad de ocurrencia

Tipos de Distribuciones en Análisis de Riesgo

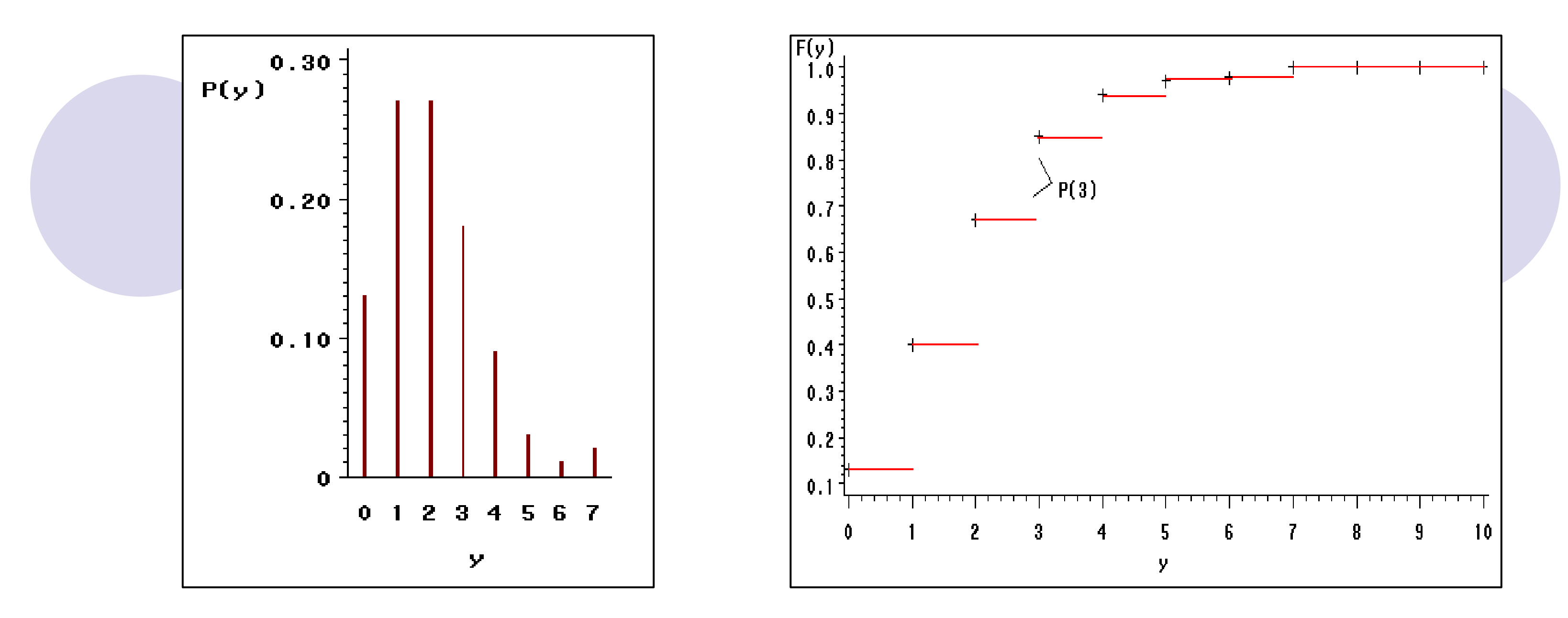
- Discretas y continuas
- Limitadas o no
- Distribuciones Empíricas
  - Existen datos
  - Utilidad para las condiciones analizadas exclusivamente
- Distribuciones obtenidas de expertos
  - Opinión obtenida de uno o más expertos
  - No reflejan evidencia empírica sino inferencias individuales o grupales
  - Pueden modificarse con información posterior

Distribuciones obtenidas de expertos

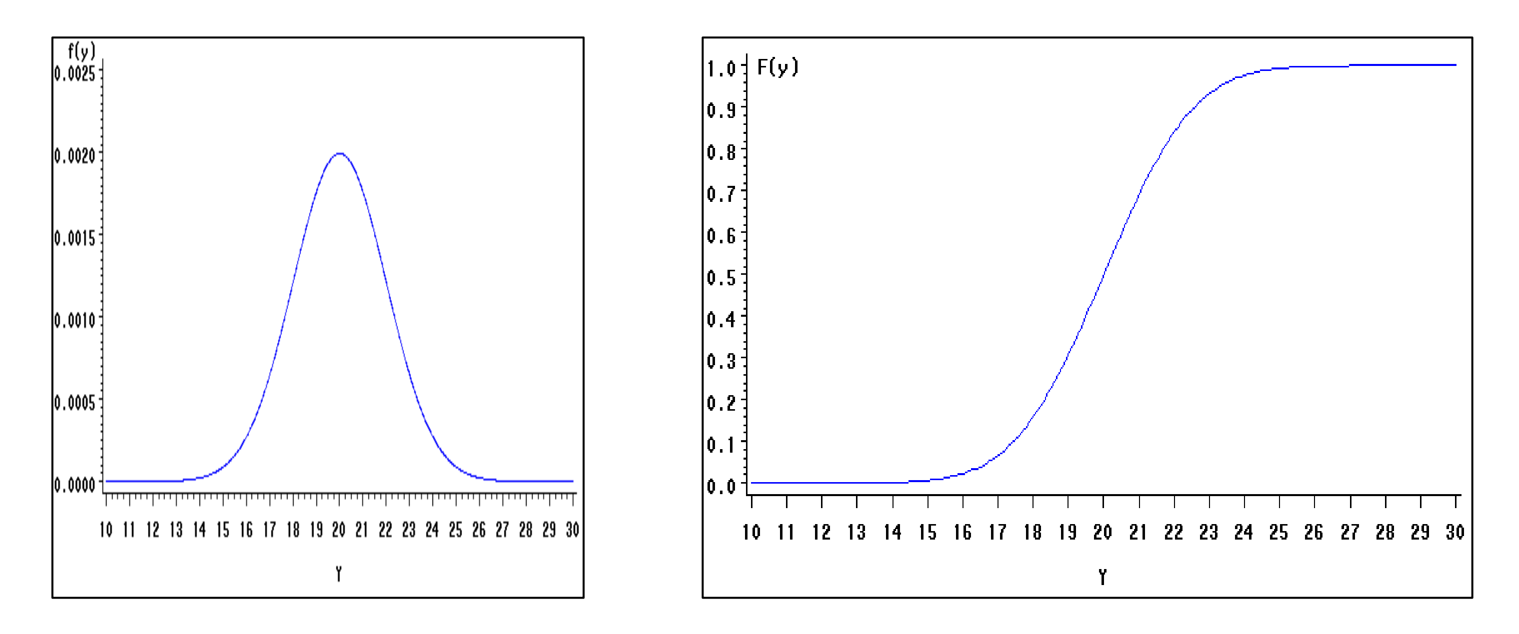
- Un *experto* es una persona con conocimiento o habilidad especial en un campo específico
- Un *dictamen* se refiere a inferencias hechas para formar una opinión
- Entonces, un *dictamen de expertos* debe ser la opinión inferencial de un especialista con respecto a un tema dentro de su especialidad

Cuándo se deriva una distribución de probabilidad a partir de expertos?

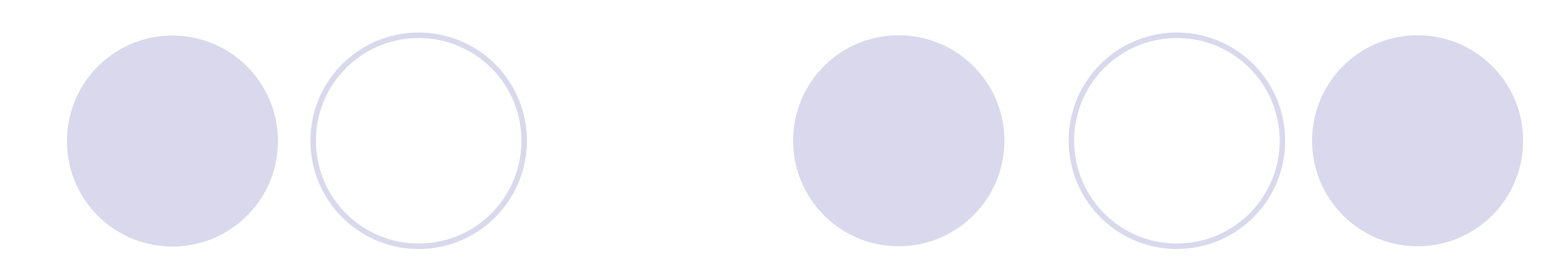
- No es posible caracterizar los parámetros a partir de los datos, debido a que:
  - *No se han recolectado datos en el pasado*
  - *Es muy dificultosa su recolección*
  - *El área de interés es novedosa*
  - *Los datos pasados no son útiles*
- En estos casos hay que recurrir a la solicitud de dictamen de expertos a través de *protocolos de elicitación*
- Se busca minimizar sesgos y ayudar al experto a construir la distribución de probabilidad subjetiva



Funciones de probabilidad y de distribución acumulada para una v.a. discreta



Funciones de densidad y de distribución acumulada para una v.a. continua



- Se representa gráficamente de dos maneras
- Ambas representan la misma distribución
- Transmiten diferente información
- Densidad:
  - La probabilidad relativa de cada valor
  - Los valores más probables (moda)
  - La forma de la distribución
- Distribución acumulada:
  - Percentiles
  - Rangos entre percentiles

Distribuciones discretas

- X ➔ refiere a la v.a. y x ➔ valor específico que puede tomar
- Función masa de probabilidad o distribución de probabilidad
- $f(x) = P(X = x)$ para todo x perteneciente al intervalo
- $f(x) \geq 0$ y $\sum f(x) = 1$

Distribuciones discretas

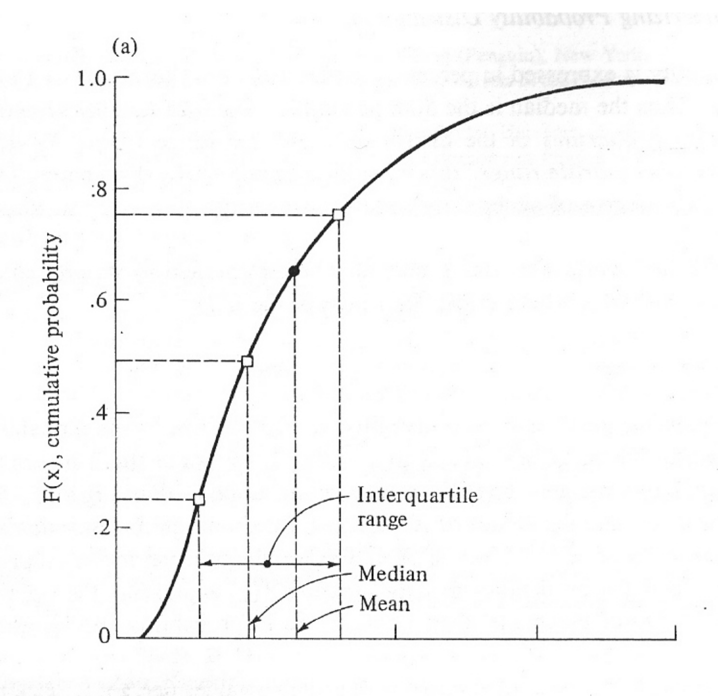
- <u>Función de distribución o Distribución acumulada</u>
- $F(x) = P(X \leq x)$
- Probabilidad acumulada de que X tome cualquier valor menor o igual a x
- En tanto *x* se incrementa de su menor a mayor valor, F(*x*) se incrementa monotónicamente, es decir,tiene pendiente no negativa
- Como representa una probabilidad, debe ir de 0 a 1 en tanto *x* va de su menor a mayor valor

Distribuciones continuas

- Función de densidad (pdf)
- Las áreas debajo de la curva dan las probabilidades relacionadas con los intervalos en el eje horizontal
- Una pdf integrada de *a* a *b* da la probabilidad de que la v.a. correspondiente tomará un valor del intervalo
- $f(x)$ pdf de $x \leftrightarrow P(a \leq x \leq b) = \int f(x)dx$, $a<b$

Distribuciones continuas

- Función de distribución acumulada (cdf)
- Al igual que en el caso discreto, en Análisis de Riesgo interesa conocer la probabilidad de que un valor de la v.a. continua sea menor o igual a un número real *x*
- $F(X) = P\,(X \leq x) = \int f(t)dt$



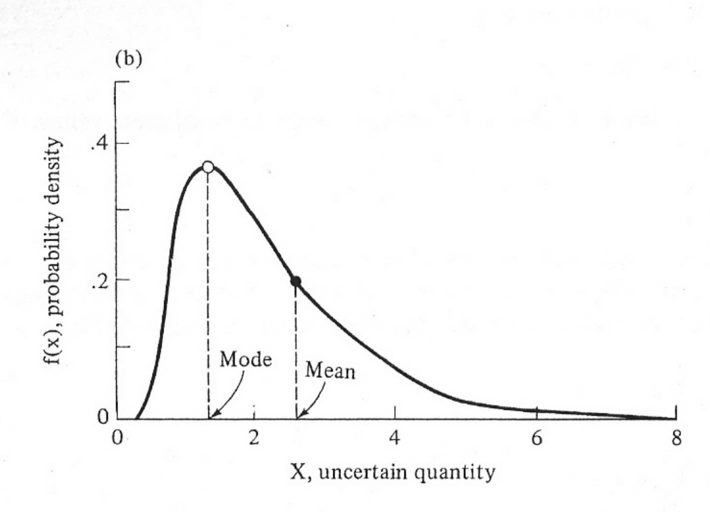


Figure 5.1. (a) A cumulative distribution function (CDF), with median and quartiles marked, above (b) the corresponding probability density function (PDF), with mode marked. Also shown is the mean. Although in this example the mean lies inside the interquartile range, for highly skewed distributions the mean may lie well outside this range.

Cuantiles de una distribución *(redefiniendo mediana y percentiles)*

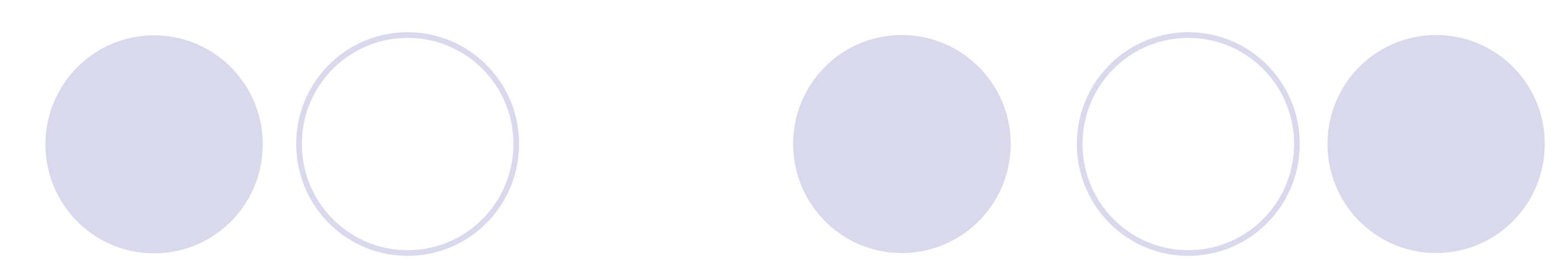
- La mediana de una v.a. es un valor tal que hay una probabilidad de 0,5 de que el verdadero valor de la v.a. sea menor a él
- Mediana = $x_{0,5}$; $P(X \leq x_{0,5}) = 0,5$
- La mediana es el cuantil 0,5 de la distribución
- Los cuantiles 0,25 y 0,75 son los cuartiles
- $[x_{0,25}; x_{0,75}]$= rango intercuartil

Cuantiles de una distribución *(redefiniendo mediana y percentiles)*

- En general el *cuantil p* de una distribución es un valor tal que hay una probabilidad *p* de que el verdadero valor de la variable sea menor a él
- Cualquier valor x / P ($X \leq x$) = p
- Si expresamos la probabilidad en porcentaje, entonces el valor cuantil se refiere como *percentil*
- Percentil 99 = x / P ($X \leq$ x) = 0,99
- Los percentiles pueden leerse directamente del formato CDF, pero no de un gráfico de PDF

Momentos de una distribución

- Primer momento: media o valor esperado, E(X)
- Segundo momento: Varianza, σ^2
- Tercer momento: Sesgo
- Cuarto momento: Curtosis



Sesgo

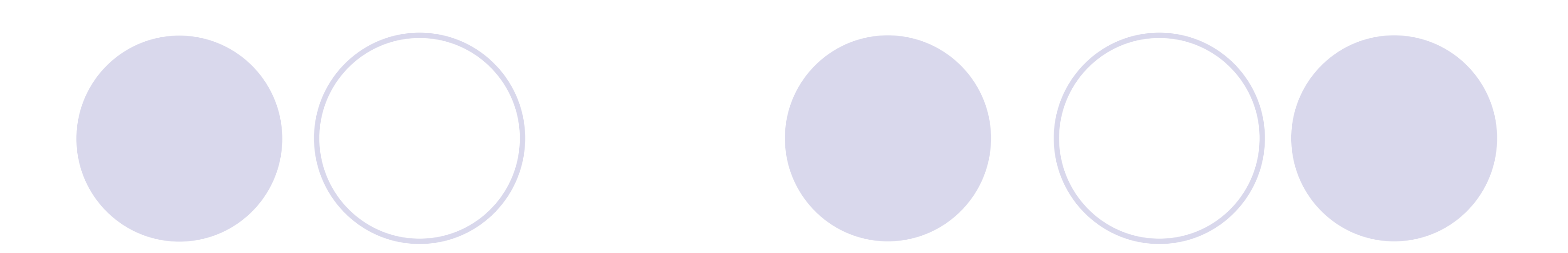
Positivo: las frecuencias más altas se encuentran a la izquierda del valor central, mientras que a la derecha hay frecuencias más pequeñas (*cola*).

Negativo: Cuando la cola está en el lado izquierdo.

sesgo +
(cola a derecha)

sesgo -
(cola a izquierda)

sesgo 0
simetría



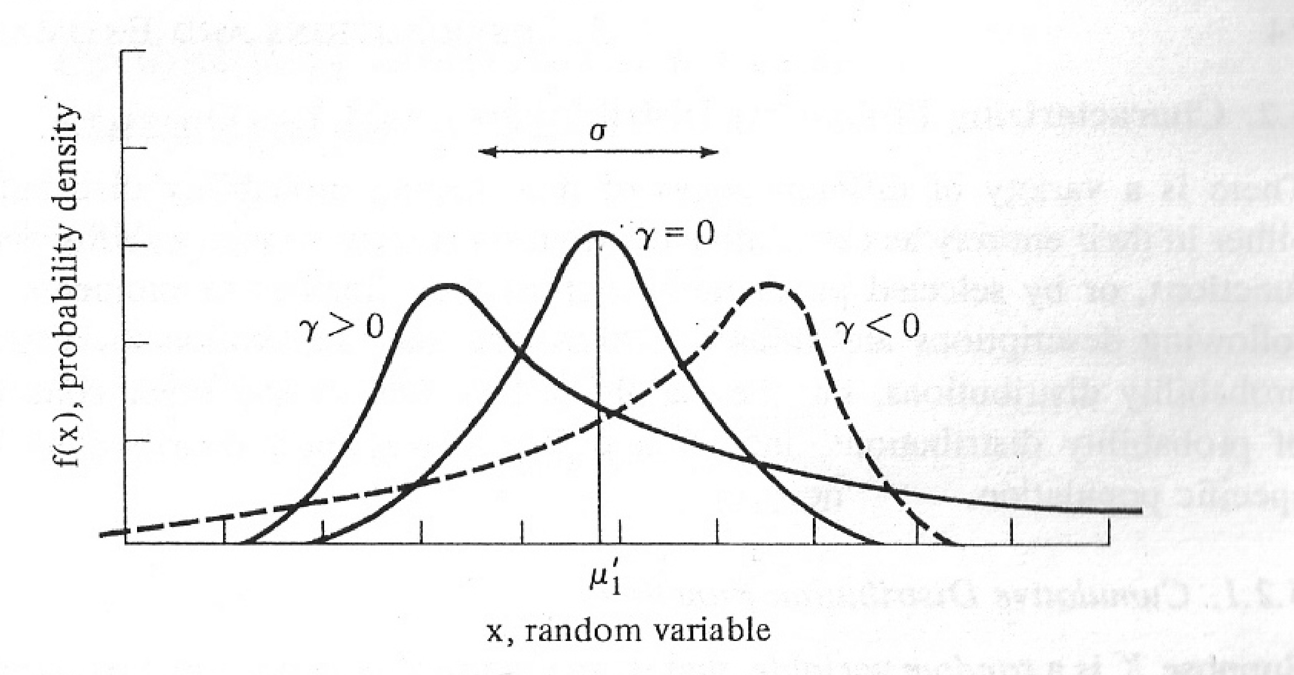


Figure 5.2. Probability density functions with negative, zero, and positive coefficients of skewness.

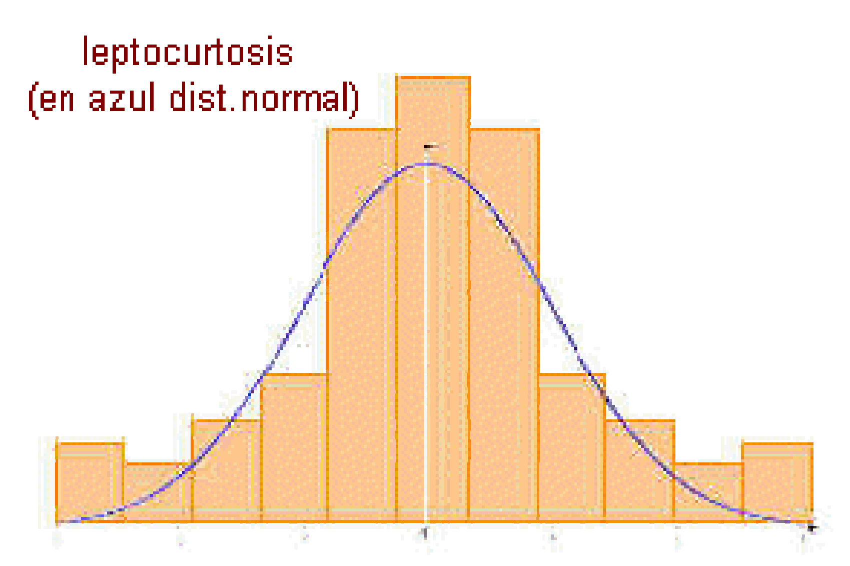
Curtosis

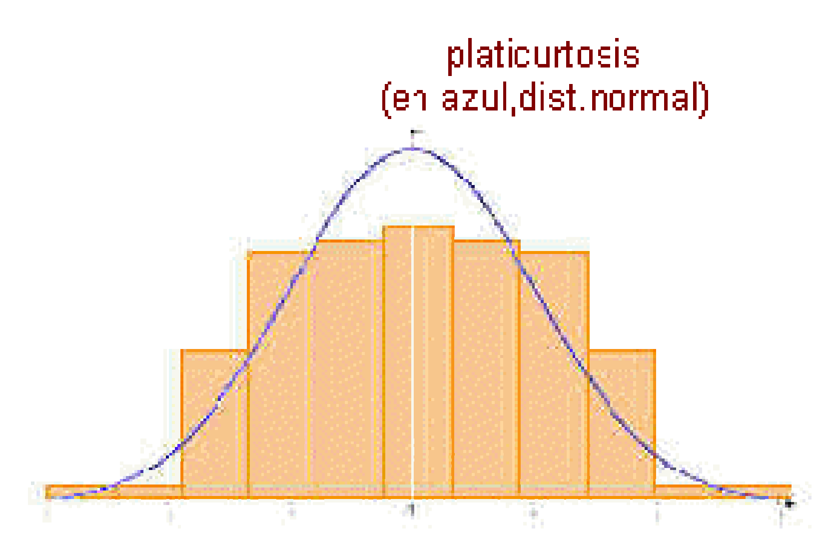
El grado de "achatamiento" se compara con la distribución ***normal***

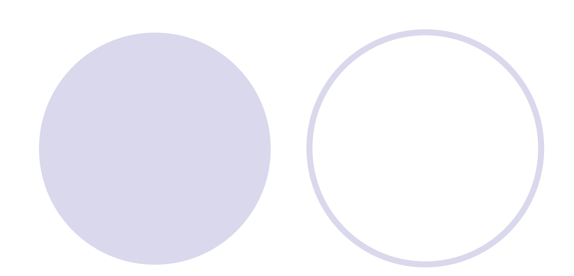
Leptocúrtica: la distribución de frecuencias es más "picuda" que la normal, K > 0

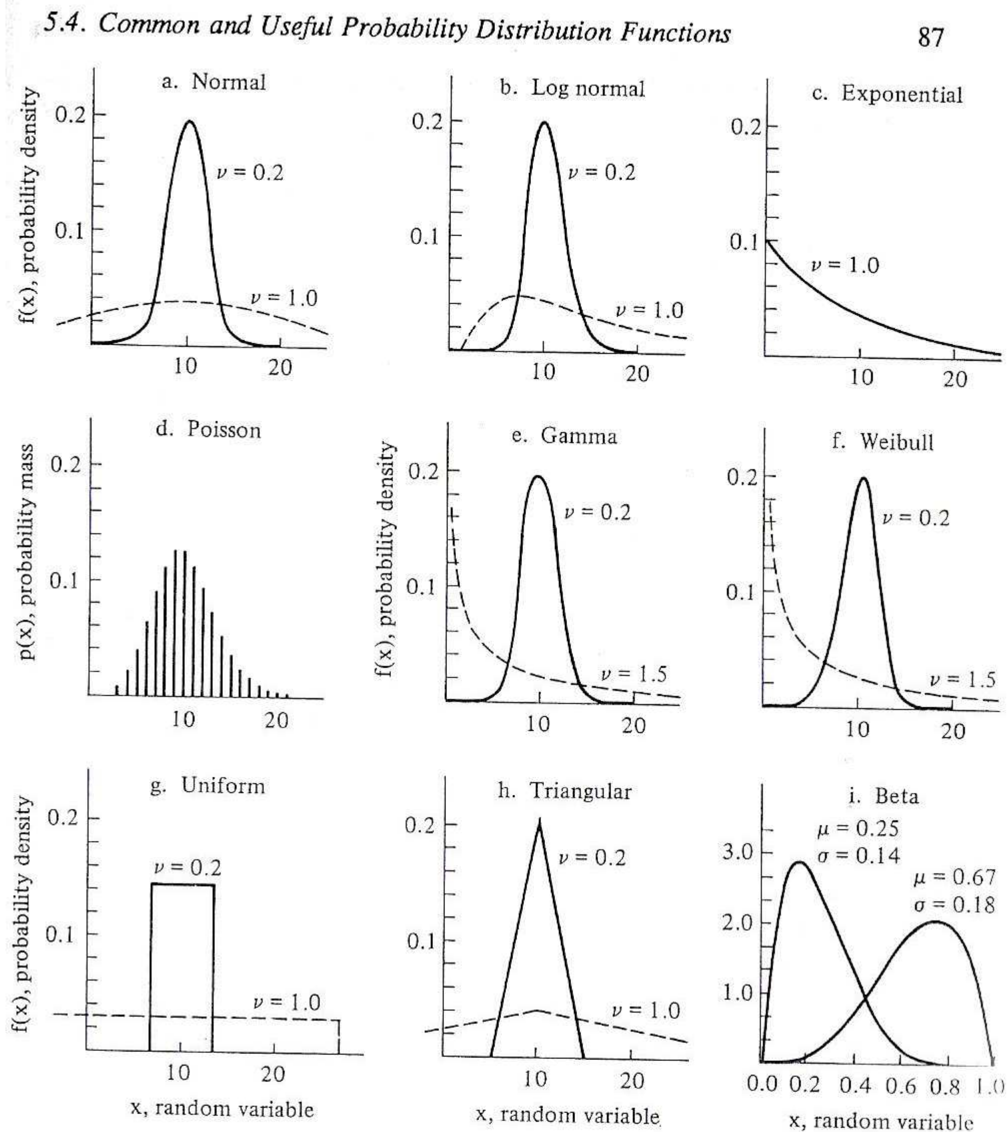
Mesocúrtica: la distribución de frecuencias es tan "picuda" como la normal, K = 0

Platicúrtica: la distribución de frecuencias es más "achatada" que la normal, K < 0









5.4. *Common and Useful Probability Distribution Functions* 87

Figure 5.4. Plots of common probability distributions discussed in the text: a. Normal, b. Lognormal, c. Exponential, d. Poisson, e. Gamma, f. Weibull, g. Uniform, h. Triangular, i. Beta.

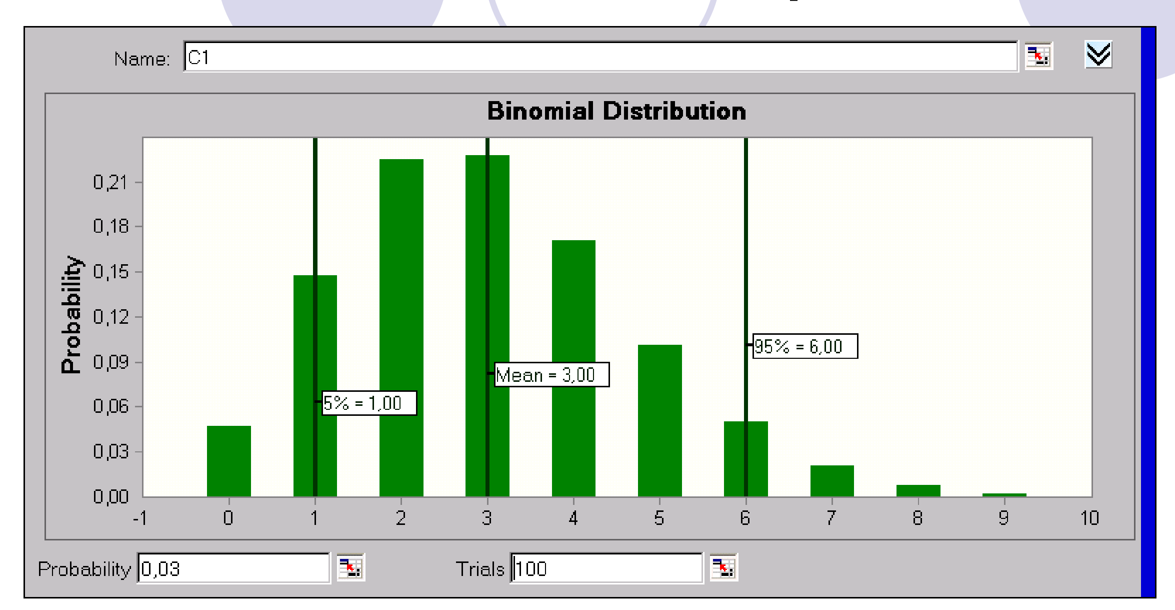
Gráficas de pdf o función masa, con E(x)= 10 y v= coef. variación Fuente Morgan y Henrion, 1990

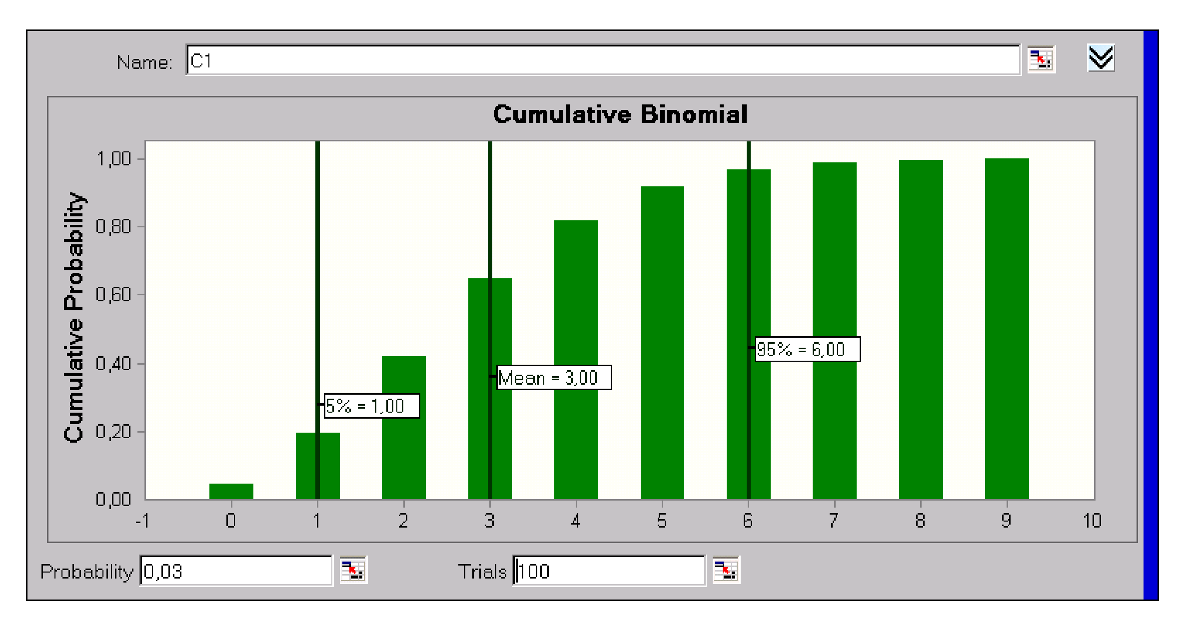
Ejemplos de utilización de distribuciones de probabilidad en riesgo ambiental

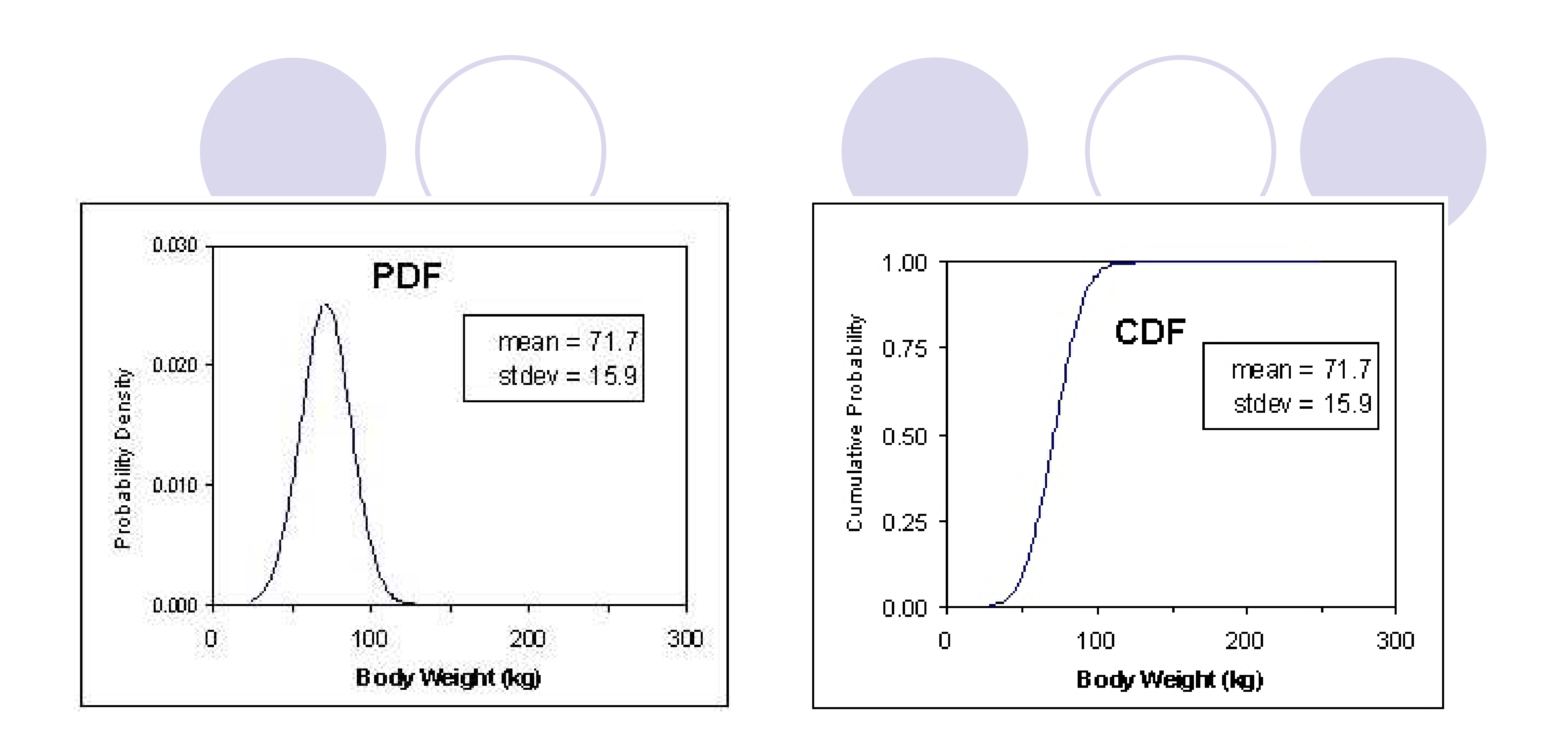
- *Las distribuciones de probabilidad a menudo surgen de las propiedades fundamentales de las variables que intentamos representar*

Distribución de gansos enfermos

Prevalencia en la población 3%





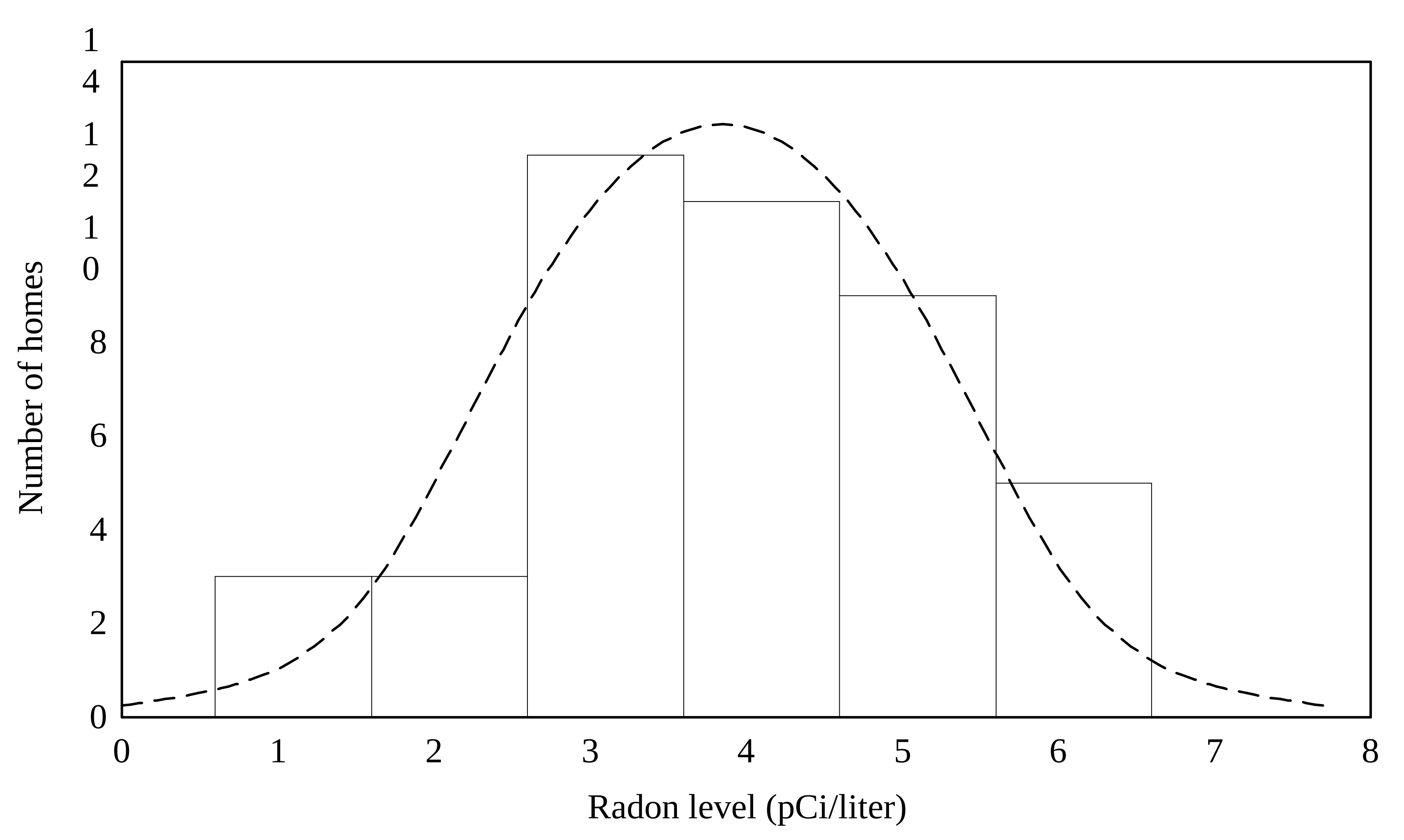


Distribución normal que caracteriza la variabilidad en peso corporal de adultos (varones y mujeres). Fuente: Finley y Paustenbach, 1994

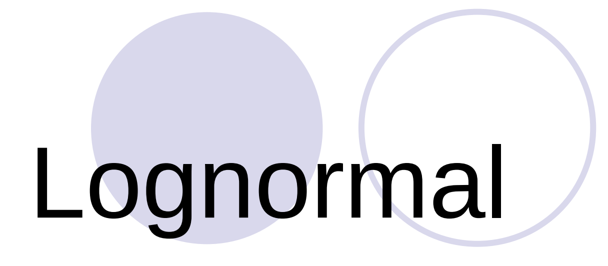
Estudio de Caso

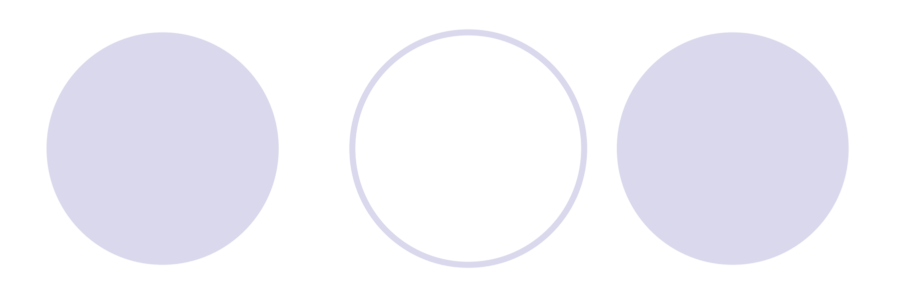
- El Radón se encuentra en el interior de las viviendas en todo el país
- No es posible medir todas las viviendas pero se tiene una muestra de más de 1000
- Importa cómo se modela la distribución de la concentración de radón

Distribución Normal (Muestra de 43 casas)



Kammen y Hassenzahl, 1999



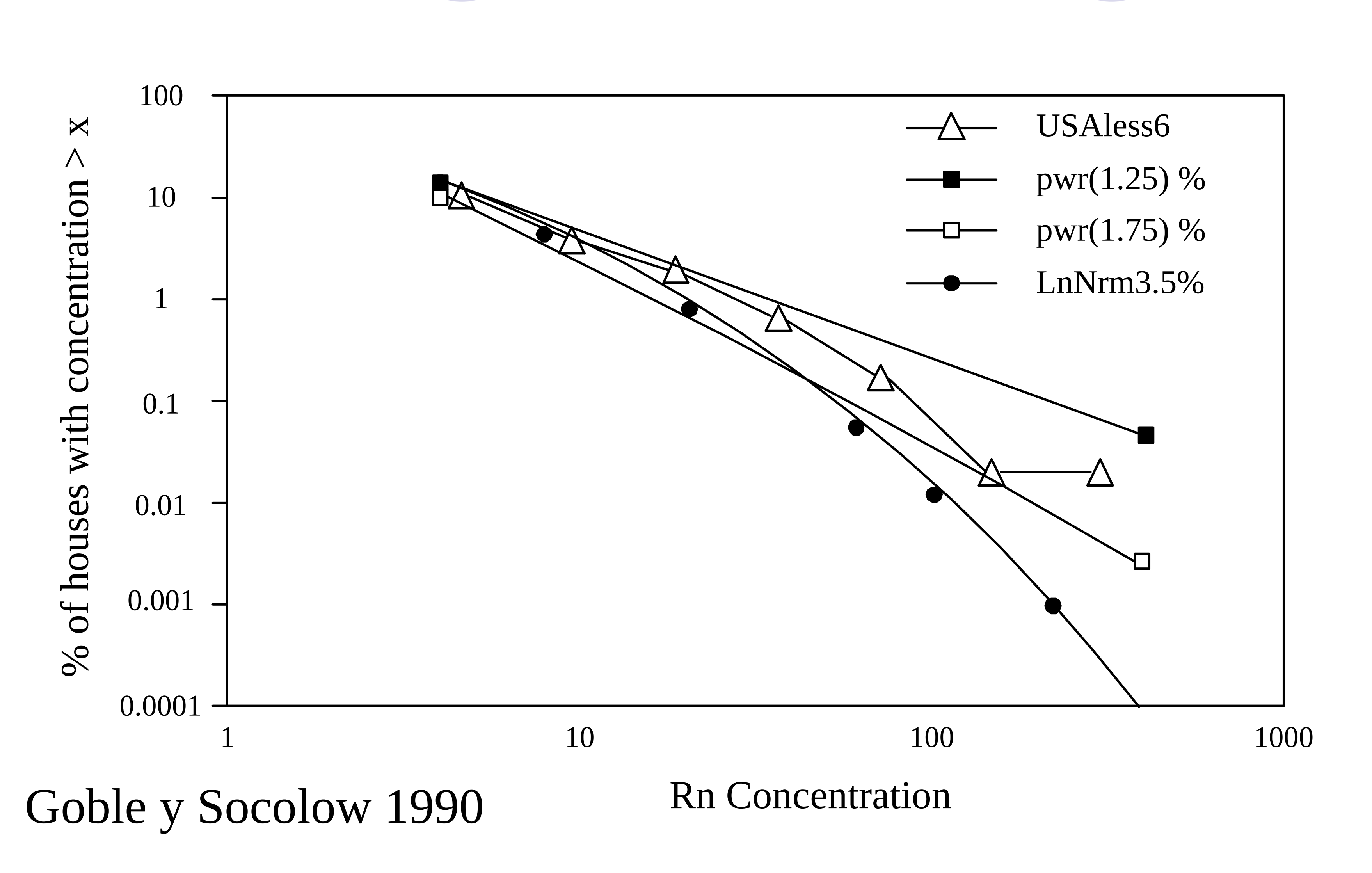


Lognormal

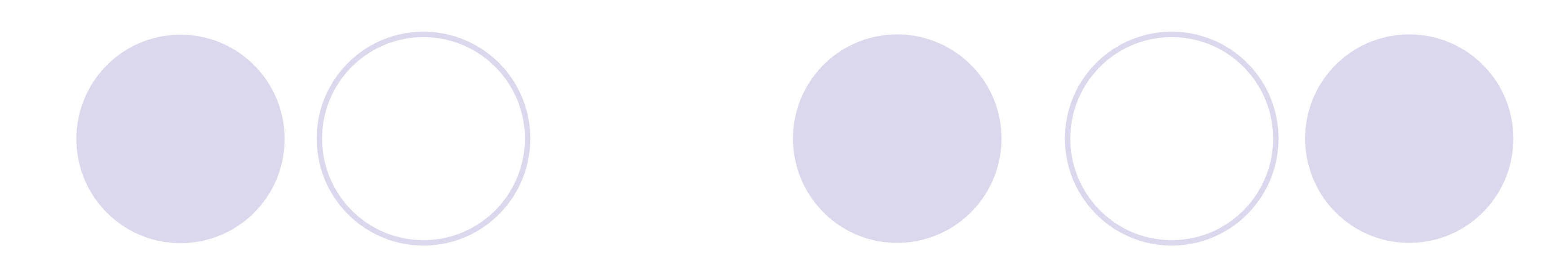
20
15
10
5
0
Houses (%)
>8
0
2
4
6
8
^{222}Rn (pCi/liter)

Nero et al 1987

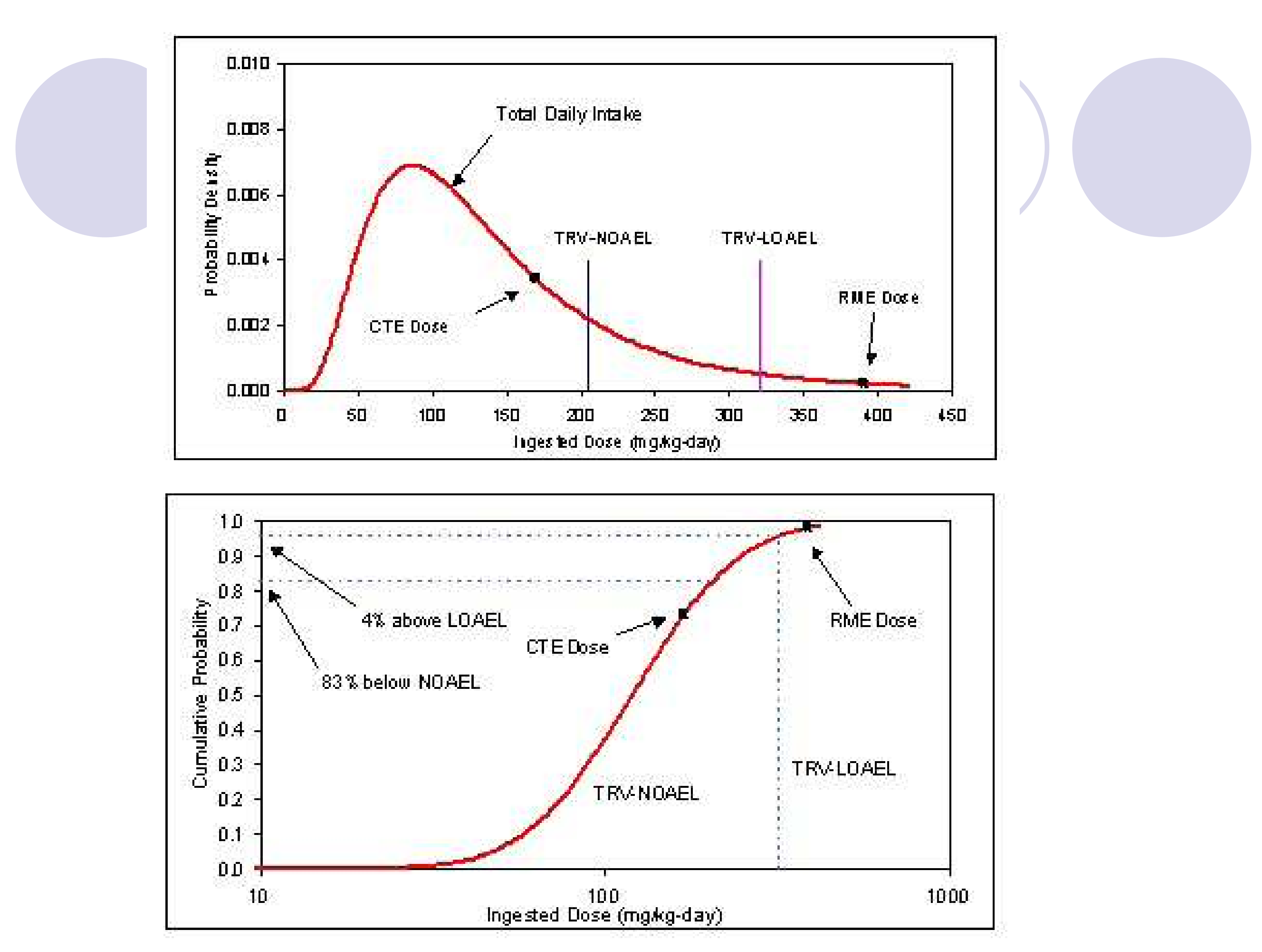
Exponencial vs. Lognormal



Goble y Socolow 1990

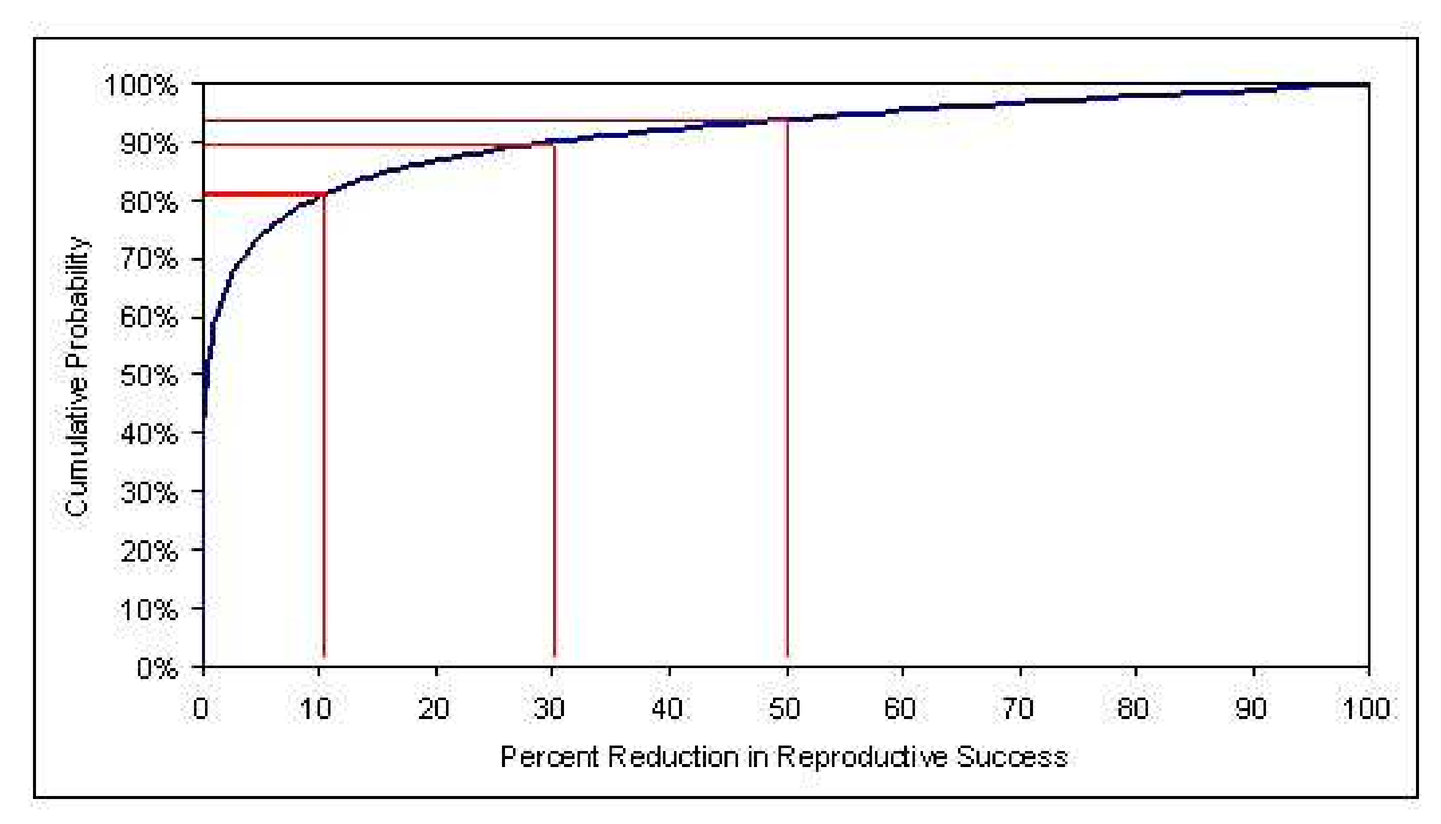


- Los gráficos de distribución acumulada son útiles para encontrar *intervalos de probabilidad*
- La probabilidad entre dos valores, es la diferencia entre sus probabilidades acumuladas
- Proporción de la población por debajo o por encima de un valor crítico
- Rangos entre percentiles:
- P95 – P5 = 90% central de los valores

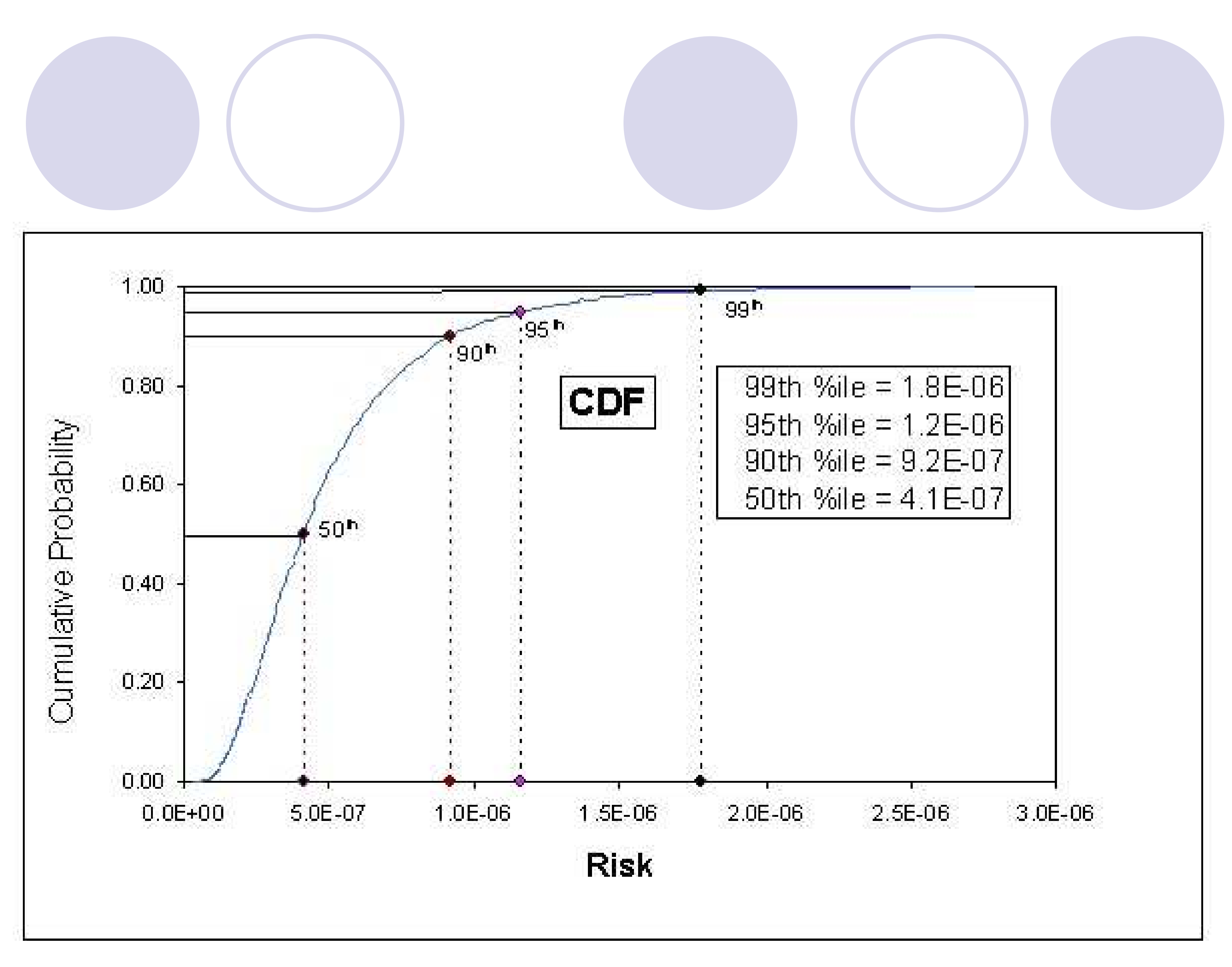


Evaluación de la fracción de la población por debajo y por encima de valores de toxicidad de referencia. Fuente: EPA, 2001.

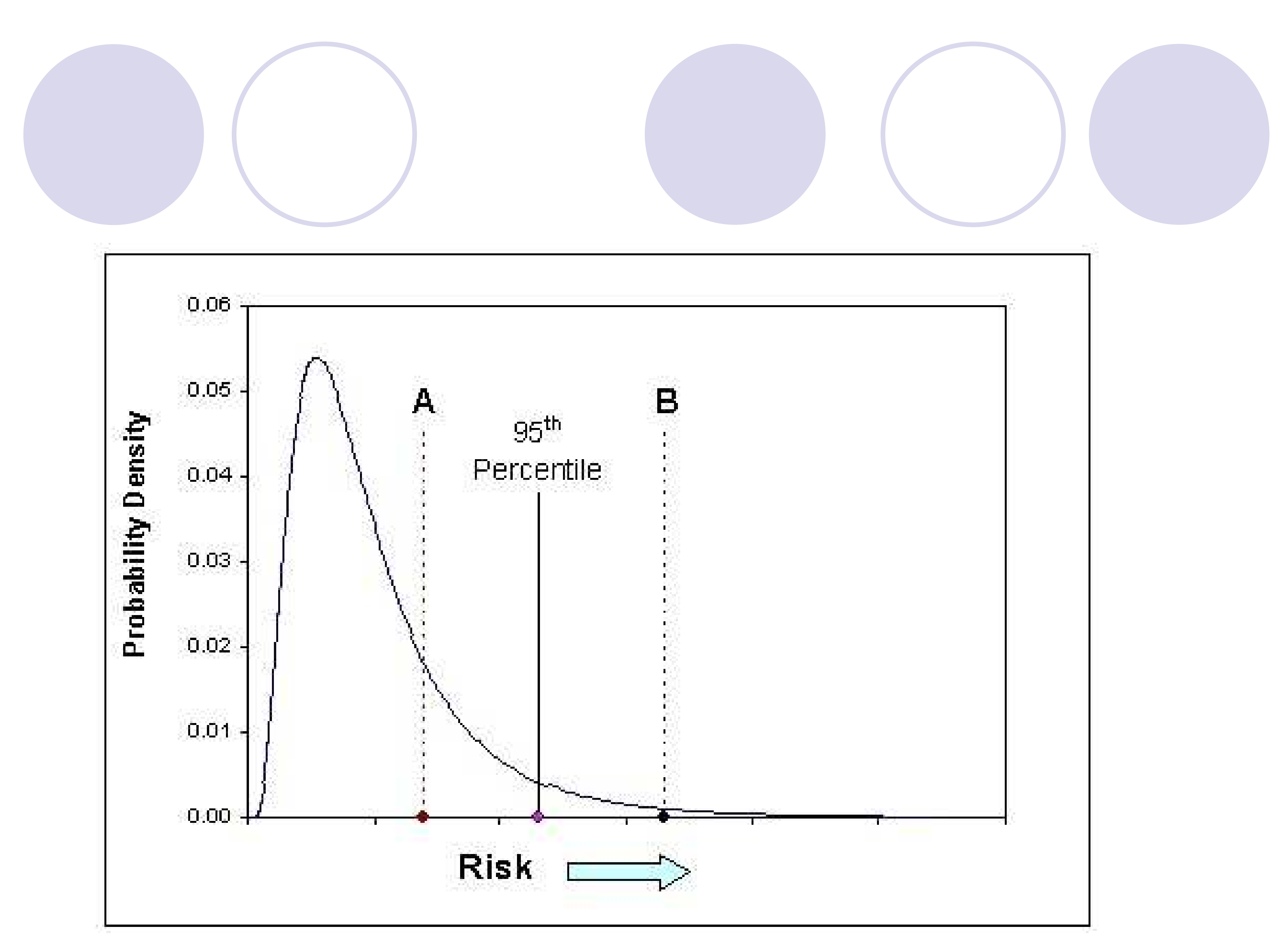
x	Total daily intake
a	0
b	100
k	0.9
n	2.5
alpha	1.6
rho	1.3



Percent Reduction	Percent of Population
0-10%	81%
10-30%	9%
30-50%	4%
>50%	6%



Resultados de un análisis probabilístico mostrando la función de distribución acumulada del riesgo incremental de cáncer de por vida



Distribución de probabilidad del riesgo ilustrando el percentil 95 y dos posibles niveles críticos, A y B. En el caso A se debe garantizar una acción de remediación. En el caso B no sería necesario.